the diseased animal creates an *Nsi*I site in the control sequence (Figure 7), and therefore digestion with this enzyme distinguishes between them, producing a 428 bp fragment from the control (lane 1 in Figure 7) and a 492 bp fragment from the sheep with ceroid lipofuscinosis (lane 16 in Figure 7). Some animals known to have ceroid lipofuscinosis were found that were homozygous for both of the two previously determined pseudogene sequences, and others were heterozygous for these genes. Therefore there is no linkage between the expressed P2 pseudogene and the gene causing ovine ceroid lipofuscinosis.

Possible causes of ceroid lipofuscinosis

After these investigations, the most plausible hypothesis for the basis of the ceroid lipofuscinosis is that the disease arises from a lesion in the degradative pathway of subunit c. It is thought that lysosomes digest mitochondria by a process of autophagocytosis (Pfeifer, 1987), but turnover rates of proteins of the mitochondrial inner membrane differ widely (Hare and Hodges, 1982), implying that selective turnover mechanisms may operate within the mitochondrion. However, the turnover rate of subunit c has not been measured, and it is not known whether mitochondrial proteases are involved in its degradation, nor if its degradation takes place in lysosomes.

In seeking the basis of ceroid lipofuscinosis, the distinctive properties of the c subunit of ATP synthase should also be borne in mind. First, ATP synthase is an abundant component of mitochondrial membranes, and probably has 10–12 copies of the c subunit per enzyme complex. Therefore, on a molar basis, the c subunit is one of the most abundant proteins in the inner membrane. It is also one of the most hydrophobic proteins that has been studied, and it falls into the category of proteolipid (a protein with the solubility properties of a lipid), being readily soluble in chloroform/methanol mixtures and insoluble in aqueous solutions. Therefore any mechanism for degrading the subunit would have to be able to cope with these physicochemical properties. However, subunit c is by no means the only proteolipid in mitochondrial inner membranes, and at least 15 different proteins have been found in chloroform/methanol extracts of mitochondria (Fearnley and Walker, 1986, 1987). None of these other mitochondrial proteolipids are found in the storage bodies associated with ceroid lipofuscinosis. This observation may indicate that specific factors operate in the degradation and turnover of the c subunit.

REFERENCES

Bentley, D. L. and Rabbitts, T. H. (1981) Cell **24**, 613–623
Benton, W. D. and Davis, R. W. (1977) Science **196**, 180–182
Boustany, R. N., Alroy, J. and Kolodny, E. H. (1988) Am. J. Med. Gen. Suppl. **5**, 47–58
Chirgwin, J. M., Przbyla, A. F., MacDonald, A. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Dyer, M. R. and Walker, J. E. (1993) Biochem. J. **293**, 51–64
Dyer, M. R., Gay, N. J. and Walker, J. E. (1989) Biochem. J. **260**, 249–258
Dyken, P. R., (1988) Am. J. Med. Gen. Suppl. **5**, 69–84
Farrell, P. J., Deininger, P. L., Bankier, A. and Barrell, B. G. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 1565–1569
Fearnley, I. M. and Walker, J. E. (1986) EMBO J. **5**, 2003–2008
Fearnley, I. M. and Walker, J. E. (1987) Biochemistry **26**, 8247–8251
Fearnley, I. M., Walker, J. E., Martinus, R. D., Jolly, R. D., Kirkland, K. B., Shaw, G. J. and Palmer, D. N. (1990) Biochem. J. **268**, 751–758
Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13
Gardiner, R. M. (1992) Am. J. Hum. Genet. **42**, 539–541
Gay, N. J. and Walker, J. E. (1985) EMBO J. **4**, 3519–3524
Hare, J. F. and Hodges, R. (1982) J. Biol. Chem. **257**, 3575–3580
Jolly, R. D., Janmaat, A., West, D. M. and Morrison, I. (1980) Neuropathol. Appl. Neurobiol. **6**, 195–209
Jolly, R. D., Shimada, A., Dopfemer, I., Slack, P. M., Birtles, M. J. and Palmer, D. N. (1989) Neuropathol. Appl. Neurobiol. **15**, 371–383
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Messing, J. (1983) Methods Enzymol. **101**, 20–78
Messing, J. and Vieira, J. (1982) Gene **19**, 269–276
Norrander, J., Kempe, T. and Messing, J. (1983) Gene **26**, 101–106
Palmer, D. N., Barns, G., Husbands, D. R. and Jolly, R. D. (1986) J. Biol. Chem. **261**, 1773–1777
Palmer, D. N., Martinus, R. D., Cooper, S. M., Midwinter, G. G., Reid, J. C. and Jolly, R. D. (1989) J. Biol. Chem. **264**, 5736–5740
Palmer, D. N., Fearnley, I. M., Walker, J. E., Hall, N. A., Lake, B. D., Wolfe, L. S., Haltia, M., Martinus, R. D. and Jolly, R. D. (1992) Am. J. Med. Genet. **41**, 561–567
Pfeifer, U. (1987) in Lysosomes: Their Role in Protein Breakdown, pp. 3–59, Academic Press, London
Proudfoot, N. J. and Brownlee, G. G. (1976) Nature (London) **263**, 211–214
Rider, J. A. and Rider, D. L. (1988) Am. J. Med. Gen. Suppl. **5**, 21–26
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scarf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Ehrlich, H. A. (1988) Science **239**, 487–491
Sebald, W. and Hoppe, J. (1981) Curr. Top. Bioenerget. **12**, 1–64
Southern, E. M. (1975) J. Mol. Biol. **98**, 503–517
Viñas, O., Powell, S. J., Runswick, M. J., Iacobazzi, V. and Walker, J. E. (1990) Biochem. J. **265**, 321–326
Walker, J. E., Gay, N. J., Powell, S. J., Kostina, M. and Dyer, M. R. (1987) Biochemistry **26**, 8613–8619
Walker, J. E., Arizmendi, J. M., Dupuis, A., Fearnley, I. M., Finel, M., Medd, S. M., Pilkington, S. J., Runswick, M. J. and Skehel, J. M. (1992) J. Mol. Biol. **226**, 1051–1072
Zeman, W. and Dyken, P. R. (1969) Pediatrics **44**, 570–583

APPENDIX

Evolution of the expressed P2 pseudogene and the origin of the P1 and P2 genes

Simon M. MEDD and John E. WALKER

The approximate elapsed time (T_0) since the divergence of two related sequences was calculated according to Miyata and Yasunaga (1981). The sequence difference (K) is defined as the number of mismatches per nucleotide site of two aligned sequences. It is assumed that nucleotide substitution follows a Poisson distribution, that it is equiprobable for all four nucleotides, and that multiple and back mutations may take place at each site. The rate of substitution (v) per nucleotide site per year is given by:

$$2vT_0 = -\tfrac{3}{4}\ln[1-\tfrac{4}{3}K], \text{ where } K < \tfrac{3}{4}.$$

Estimation of the ages of the pseudogenes

From fossil evidence it is thought that humans diverged from cattle and sheep approx. 68 million years ago, and that sheep and cattle diverged from each other approx. 50 million years later (Romero-Herrera et al., 1978). Using these T_0 values, and the cDNA sequence differences between humans and cattle, humans and sheep, and cattle and sheep for both P1 and P2 cDNAs (Table A1) evolutionary rates (v) for the P1 and P2 precursors were calculated (Table A2). These calculations showed that the rate of fixation of mutations in both P1 and P2 is slower between

Table A1 Comparison of nucleotide sequences of human, bovine and ovine P1 and P2 cDNAs

H, B and O refer to human, bovine and ovine sequences respectively. *K*[H–B] etc. are the sequence differences between corresponding sequences in the two specified species. The entire cDNA is the region over which the two sequences can be aligned. The pre-sequences encode the mitochondrial import pre-sequences present in P1 and P2, and the mature region codes for the mature subunit c of ATP synthase. The untranslated region is the sum of the 5′ and 3′ untranslated regions in the cDNAs.

	P1			P2		
Region	*K*[H–B]	*K*[H–O]	*K*[B–O]	*K*[H–B]	*K*[H–O]	*K*[B–O]
Entire cDNA	0.121	0.109	0.016	0.136	0.130	0.020
Pre-sequence	0.137	0.143	0.016	0.228	0.223	0.029
Mature	0.044	0.044	0.004	0.052	0.052	0.000
Untranslated	0.217	0.235	0.047	0.178	0.164	0.035

Table A2 Evolutionary rates for P1 and P2 precursors derived from the sequence differences between human, bovine and ovine cDNA sequences

The evolutionary rate per nucleotide site per year (v) has been multiplied by 10^9 to give the values shown. Evolutionary rates between humans and cattle (v[H–B]) and between humans and sheep (v[H–O]) were calculated assuming that humans diverged from sheep and cattle approx. 68 million years ago, and evolutionary rates between cattle and sheep (v[B–O]) assume that they diverged 18 million years ago. See Table A1 for definitions of regions.

	P1			P2		
Region	v[H–B]	v[H–O]	v[B–O]	v[H–B]	v[H–O]	v[B–O]
Entire cDNA	0.97	0.86	0.44	1.11	1.04	0.55
Pre-sequence	1.11	1.16	0.44	1.99	1.94	0.83
Mature	0.33	0.33	0.11	0.37	0.37	0.00
Untranslated	1.88	2.07	1.33	1.49	1.36	1.00

Table A3 Comparison of the nucleotide sequences of the P1 and P2 cDNAs in man, cattle and sheep with the cDNA sequences of the ovine and human P2 pseudogenes

The notations are explained in the legend to Table A2; [Oϕ] and [Hϕ] refer to the ovine and human pseudogenes.

	P1–P2			P2–P2	
Region	*K*[H]	*K*[B]	*K*[O]	*K*[Oϕ]	*K*[Hϕ]
Entire cDNA	0.445	0.402	0.419	0.009	0.101
Pre-sequence	0.640	0.581	0.605	0.015	0.085
Mature	0.093	0.119	0.119	0.004	0.066
Untranslated	0.740	0.621	0.708	0.005	0.105

sheep and cattle than it is between humans and either sheep or cattle. It has been observed previously that the rate of fixation of mutations in the myoglobin molecule fluctuates along different lines of descent (Romero-Herrera et al., 1978). Furthermore, the rate of fixation of mutations between sheep and cows for the entire P2 cDNA is greater than that for the P1 cDNA. Therefore it seemed prudent to use the evolutionary rate of P2 between sheep and cows (0.55×10^{-9} per nucleotide site per year; see Table A2) to calculate the age of the P2-related pseudogene in sheep, using the sequence difference between the ovine P2 cDNA and the ovine P2-related pseudogene from sheep with ceroid lipofuscinosis. This gave an age for the pseudogene of approx. 8 million years.

However, the evolutionary rate used in this calculation was derived from the sequence difference between the entire cDNA sequences of the functional P2 precursors. These sequences are largely protein coding, and therefore the rates of fixation of mutations are taken from DNA sequences that are subject to evolutionary constraint. If the expressed pseudogene were non-functional, then its DNA sequence would not be subject to such constraints, and consequently its rate of evolution could be significantly higher. So, if the pseudogene were non-functional for all or most of the time since it arose, it may be considerably younger than 8 million years old. Comparison of this figure with the estimated date of divergence of sheep and cattle, around 18 million years ago, leads to the conclusion that the sheep pseudogene probably arose after the divergence of sheep from cattle.

The sequence of a human P2 pseudogene has been described (Dyer and Walker, 1993), although there is no evidence that it is expressed. Like the sheep pseudogene, it codes for residues 1–31 of the import precursor of P2, a stop codon being present in the pseudogene because of a C to T mutation in the codon for Arg-32 in the expressed gene. From an average evolutionary rate of 1.08×10^{-9} per nucleotide per year for the entire P2 cDNA between humans and cattle, and humans and sheep, it was estimated that this human P2 pseudogene arose 47 million years ago. Again, the value of the evolutionary rate used in the calculation assumes that the DNA sequence is coding and is subject to evolutionary constraint. Therefore a second estimate was made using an average evolutionary rate of 1.42×10^{-9} per nucleotide per year (see Table A2), derived from the non-protein-coding regions of the human and bovine and human and ovine P2 cDNAs. This gave an age for the pseudogene of 37 million years. However, even the untranslated regions of a transcript are functional, and therefore are also subject to some evolutionary constraint. So, if the human P2-related pseudogene has been totally redundant ever since its generation, it could be even younger than this latter value. Both estimates suggest that the pseudogene arose after the divergence of sheep and cattle from humans, about 68 million years ago. Therefore the calculations confidently predict that the two related human and ovine P2 pseudogenes of known sequence arose independently of each other from their respective P2 genes. It appears that the occurrence of a stop codon in exactly the same position in the sequences of both the human and ovine pseudogenes is coincidental. Reductions in CG content in pseudogenes have been noticed previously, and the preponderance of CA or TG dinucleotides suggests that the loss of CG arose by deamination of 5-methyl-C to T (Bird et al., 1987).

Evolution of the precursors for subunit c

Separate evolutionary rates for the pre-sequence, mature protein and untranslated regions of the cDNAs for the subunit were also calculated (see Table A2). As expected, the evolutionary rates of the untranslated regions were higher than those of the entire cDNA, presumably because these sequences are not subject to such strict evolutionary constraints as coding regions. The pre-sequences have evolved several times faster than the mature protein. Import pre-sequences of nuclear-coded mitochondrial proteins vary widely in both sequence and length, although they usually have a net positive charge. It has been suggested that their secondary structures, rather than their amino acid sequences, are conserved, and are amphiphilic (von Heijne, 1986). Even the two pre-sequences of subunit c, which probably target the same protein to the mitochondrial membranes and into the same enzyme complex, differ widely in sequence, whereas the sequence of the mature subunit c is identical in man, cattle and sheep (Gay and Walker, 1985; Dyer and Walker 1993; see the main paper). Therefore the different rates of mutation between the DNA sequences coding for the pre-sequences and the mature protein of the subunit c precursors probably reflect both the strong conservation of the mature protein and the relatively weak evolutionary constraints imposed on the pre-sequences by mitochondrial import.

The origin of the P1 and P2 genes

The P1 and P2 precursors for subunit c are related in their amino acid sequence (Gay and Walker, 1985) and their gene structures are also similar (Dyer and Walker, 1993). Therefore it is likely that they arose from a common ancestral gene. In order to estimate the date of divergence of the P1 and P2 genes, it was again necessary to choose a reliable rate of evolution. Preliminary calculations showed that the pre-sequences and untranslated regions of the P1 and P2 cDNA pairs were highly diverged. The sequence difference values approached $\frac{3}{4}$ (see Table A2). Values greater than $\frac{3}{4}$ are likely to be inaccurate, as they correspond to mutation of each base on average more than once. Therefore the considerably lower values for the mature protein were more suitable for calculating evolutionary rates and the time since divergence of P1 and P2 (see Table A3). Again it was observed that the rates of mutation derived from the sequences of cattle and sheep were different from those derived from the comparison of the human and cow or human and sheep sequences. As the latter rates were derived from evolution over a much longer time, the average value, 0.35×10^{-9} per nucleotide site per year, seemed to be the most suitable for calculation of the time since divergence of the P1 and P2 genes. In this way, three times of divergence, 141, 181 and 181 million years ago respectively, were estimated from the sequence data of the three pairs of subunit c precursor from man, cattle and sheep. It has been estimated from fossils that the major orders of mammals diverged about 75 million years ago, and that mammals diverged from birds about 300 million years ago. The three dates of divergence of P1 and P2 falls between these two dates. Therefore all mammals would be expected to have inherited both P1 and P2 genes, whereas birds would be expected either to have only have a single gene for the c subunit. Alternatively, if multiple genes for the c subunit are present in avian genomes, then they have evolved independently of P1 and P2, and more recently than the divergence of birds and mammals.

Received 21 December 1992; accepted 25 January 1993

REFERENCES

Bird, A. P., Taggart, M. H., Nicholls, R. D. and Higgs, D. R. (1987) EMBO J. **6**, 999–1004
Dyer, M. R. and Walker, J. E. (1993) Biochem. J. **293**, 51–64
Gay, N. J. and Walker, J. E. (1985) EMBO J. **4**, 3519–3524
Miyata, T. and Yasunaga, T. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 450–453
Romero-Herrera, A. E., Lehmann, F., Joysey, K. A. and Friday, A. E. (1978) Philos. Trans. R. Soc. London B **283**, 61–163
von Heinje, G. (1986) EMBO J. **5**, 1335–1342